

WHAT IS CLAIMED IS:

1. An automatic speech recognition system, which recognizes speeches in acoustic signals detected by a plurality of microphones as character information, the system comprising:

5 a sound source localization module which localizes a sound direction corresponding to a specified speaker based on the acoustic signals detected by the plurality of microphones;

a feature extractor which extracts features of speech signals contained in one or more pieces of information detected by the plurality of microphones;

10 an acoustic model memory which stores direction-dependent acoustic models that are adjusted to a plurality of directions at intervals;

an acoustic model composition module which composes an acoustic model adjusted to the sound direction, which is localized by the sound source localization module, based on the direction-dependent acoustic models in the acoustic model memory, the acoustic model composition module storing the
15 acoustic model in the acoustic model memory; and

a speech recognition module which recognizes the features extracted by the feature extractor as character information using the acoustic model composed by the acoustic model composition module.

20 2. An automatic speech recognition system, which recognizes speeches of a specified speaker in acoustic signals detected by a plurality of microphones as character information, the system comprising:

25 a sound source localization module which localizes a sound direction corresponding to the specified speaker based on the acoustic signals detected by the plurality of microphones;

a sound source separation module which separates speech signals of the specified speaker from the acoustic signals based on the sound direction localized by the sound source localization module

a feature extractor which extracts features of the speech signals separated by the sound source separation module;

an acoustic model memory which stores direction-dependent acoustic models that are adjusted to a plurality of directions at intervals;

an acoustic model composition module which composes an acoustic model adjusted to the sound direction, which is localized by the sound source localization module, based on the direction-dependent acoustic models in the acoustic model memory, the acoustic model composition module storing the acoustic model in the acoustic model memory; and

a speech recognition module which recognizes the features extracted by the feature extractor as character information using the acoustic model composed by the acoustic model composition module.

3. A system according to claim 1 or 2, wherein the sound source localization module is configured to execute a process comprising:

performing a frequency analysis for the acoustic signals detected by the microphones to extract harmonic relationships;

acquiring an intensity difference and a phase difference for the harmonic relationships extracted through the plurality of microphones;

acquiring belief factors for a sound direction based on the intensity difference and the phase difference, respectively; and

determining a most probable sound direction.

4. A system according to any one of claims 1 to 3, wherein the sound source localization module employs scattering theory that generates a model for an acoustic signal, which scatters on a surface of a member to which the microphones are attached, according to a sound direction so as to specify the sound direction for the speaker with the intensity difference and the phase difference detected from the plurality of microphones.

5. A system according to any one of claims 2 to 4, wherein the sound source separation module employs an active direction-pass filter so as to separate speeches, the filter being configured to execute a process comprising:

separating speeches by a narrower directional band when a sound direction, which is localized by the sound source localization module, lies close to a front, which is defined by an arrangement of the plurality of microphones; and

separating speeches by a wider directional band when the sound direction lies apart from the front.

6. A system according to any one of claims 1 to 5, wherein the acoustic model composition module is configured to compose an acoustic model for the sound direction by applying weighted linear summation to the direction-dependent acoustic models in the acoustic model memory, and weights introduced into the linear summation are determined by training.

7. A system according to any one of claims 1 to 6, further comprising a speaker identification module,

wherein the acoustic model memory possesses the direction-dependent

acoustic models for respective speakers, and

wherein the acoustic model composition module is configured to execute a process comprising:

referring to direction-dependent acoustic models of a speaker who is identified by the speaker identifying module and to a sound direction localized by the sound source localization module;

composing an acoustic model for the sound direction based on the direction-dependent acoustic models in the acoustic model memory; and

storing the acoustic model in the acoustic model memory.

8. An automatic speech recognition system, which recognizes speeches of a specified speaker in acoustic signals detected by a plurality of microphones as character information, the system comprising:

a sound source localization module which localizes a sound direction corresponding to the specified speaker based on the acoustic signals detected by the plurality of microphones;

a stream tracking module which stores the sound direction localized by the sound source localization module so as to estimate a direction in which the specified speaker is moving, the stream tracking module estimating a current position of the speaker according to the estimated direction;

a sound source separation module which separates speech signals of the specified speaker from the acoustic signals based on a sound direction, which is determined by the current position of the speaker estimated by the stream tracking module;

a feature extractor which extracts features of the speech signals separated by the sound source separation module;

an acoustic model memory which stores direction-dependent acoustic models that are adjusted to a plurality of directions at intervals;

an acoustic model composition module which composes an acoustic model adjusted to the sound direction, which is localized by the sound source
5 localization module, based on the direction-dependent acoustic models in the acoustic model memory, the acoustic model composition module storing the acoustic model in the acoustic model memory; and

a speech recognition module which recognizes the features extracted by the feature extractor as character information using the acoustic model
10 composed by the acoustic model composition module.